Tang, M. S., & Patrick, M. H. (1977) *Photochem. Photobiol. 26*, 247–255.
Terry, B. J., Jack, W. E., Rubin, R. A., & Modrich, P. (1983) *J. Biol. Chem. 258*, 9820–9825.
Thiagalingam, S., & Grossman, L. (1991) *J. Biol. Chem.* (in press).
Van Houten, B. (1990) *Microbiol. Rev. 54*, 18–51.
Van Houten, B., Gamper, H., Sancar, A., & Hearst, J. E. (1987) *J. Biol. Chem. 262*, 13180–13187.
Van Houten, B., Gamper, H., Hearst, J. E., & Sancar, A. (1988) *J. Biol. Chem. 263*, 16533–16560.
von Hippel, P. H., Revzin, A., Gross, C. A., & Wang, A. C. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 4808–4812.
Weast, R. C., Ed. (1971) *Handbook of Chemistry and Physics*, 51st ed., Chemical Rubber Publishing Company, Cleveland, OH.
Woodbury, C. P., Jr., & von Hippel, P. H. (1983) *Biochemistry 22*, 4730–4737.
Wu, H.-M., & Crothers, D. M. (1984) *Nature 308*, 509–513.
Yarus, M., & Berg, P. (1970) *Anal. Biochem. 35*, 450–465.
Yeung, A. T., Mattes, W. B., Oh, E. Y., & Grossman, L. (1983) *Proc. Natl. Acad. Sci. U.S.A 80*, 6157–6161.
Yeung, A. T., Mattes, W. B., & Grossman, L. (1986a) *Nucleic Acids Res. 14*, 2567–2582.
Yeung, A. T., Mattes, W. B., Oh, E. Y., Yoakum, G. H., & Grossman, L. (1986b) *Nucleic Acids Res. 14*, 8535–8556.

Crystal Structure Analysis of the B-DNA Dodecamer CGTGAATTCACG[†,‡]

Teresa A. Larsen, Mary L. Kopka, and Richard E. Dickerson*
Molecular Biology Institute, Department of Chemistry and Biochemistry, and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, Los Angeles, California 90024-1570
Received September 11, 1990; Revised Manuscript Received January 3, 1991

ABSTRACT: The crystal structure of the DNA dodecamer C-G-T-G-A-A-T-T-C-A-C-G has been determined at a resolution of 2.5 Å, with a final *R* factor of 15.8% for 1475 nonzero reflections measured at 0 °C. The structure is isomorphous with that of the Drew dodecamer, with the space group $P2_12_12_1$ and cell dimensions of $a = 24.94$ Å, $b = 40.78$ Å, and $c = 66.13$ Å. The asymmetric unit contains all 12 base pairs of the B-DNA double helix and 36 water molecules. The structure of C-G-T-G-A-A-T-T-C-A-C-G is very similar to that of C-G-C-G-A-A-T-T-C-G-C-G, with no major alterations in helix parameters. Water peaks in the refined structure appear to represent a selection of peaks that were observed in the Drew dodecamer. The minor-groove spine of hydration at 2.5 Å is fragmentary, but as Narendra et al. (1991) [*Biochemistry* (following paper in this issue)] have observed, lowering the temperature leads to a more complete representation of the spine.

This paper reports the crystal structure analysis of the B-DNA dodecamer with the sequence C-G-T-G-A-A-T-T-C-A-C-G. The sequence was chosen as a variant of the "Drew" dodecamer, C-G-C-G-A-A-T-T-C-G-C-G, as part of an investigation of lexitropsins, synthetic analogues of netropsin that are designed to "read" a particular base sequence along the floor of the minor groove of B-DNA. Netropsin (Figure 1a) is a planar drug molecule that binds within the minor groove and requires a binding region of four successive A·T base pairs (Zimmer, 1975; Kopka et al., 1985a,b). It can be regarded as a polymer whose monomer is a planar amide coupled to a five-membered pyrrole ring. Longer analogues of the natural netropsin molecule in Figure 1a with eight or nine monomer units have been synthesized by Lown, Dervan, and others (Lown & Krowicki 1985; Wade & Dervan 1987). As expected, the longer analogues demand longer stretches of A·T base pairs as binding sites. The A·T specificity arises from a combination of steric hindrance from the N2 amines of the guanines along the floor of the minor groove, the intrinsic narrowness of A·T zones of the minor groove in B-DNA (Yoon et al., 1988), and the deeper electrostatic potential in A·T regions of the minor groove than in G·C regions (Pullman, 1983).

Kopka et al. (1985a,b) and Lown et al. (1986a,b) proposed independently that netropsin be modified to make it G·C specific at a given position by replacing a pyrrole by furan or imidazole (Figure 1b). This replacement of a ring C-H by O or N would simultaneously make room for the N2 amine group of guanine and provide a new hydrogen-bond acceptor that could bond to that amine. Hence in a long synthetic -(amide-ring)$_n$- analogue of netropsin, each pyrrole-ring site would demand an A·T base pair, and each imidazole-ring site might require a G·C pair. In this manner one might be able to engineer a drug analogue capable of "reading" a specific sequence of the order of 10 base pairs in length. Such a sequence, chosen because it was known to be present in a tumor or foreign cell, would occur at random in the host genome only once in $2^{10} = 1024$ times, assuming that a lexitropsin site could distinguish between an A·T and a G·C pair but could not detect the orientation of the pair (Dickerson et al., 1986). By directing or vectoring the groove-binding drug specifically to the aberrant cell, one might hope for a substantial reduction in toxicity of the drug and of unwanted chemotherapeutic side effects.

Among the first lexitropsins to be synthesized by Lown were analogues of netropsin with one or the other of the pyrroles replace by imidazole. Such a molecule should recognize a four base pair site in which one of the two inner base pairs was G·C

[†] This research was supported by NSF Grant DNB85-01682 and NIH Program Project Grant GM-31299.
* To whom correspondence should be addressed, at the Molecular Biology Institute.
[‡] Crystallographic coordinates and structure factors have been submitted to the Brookhaven Protein Data Bank.

0006-2960/91/0430-4443$02.50/0

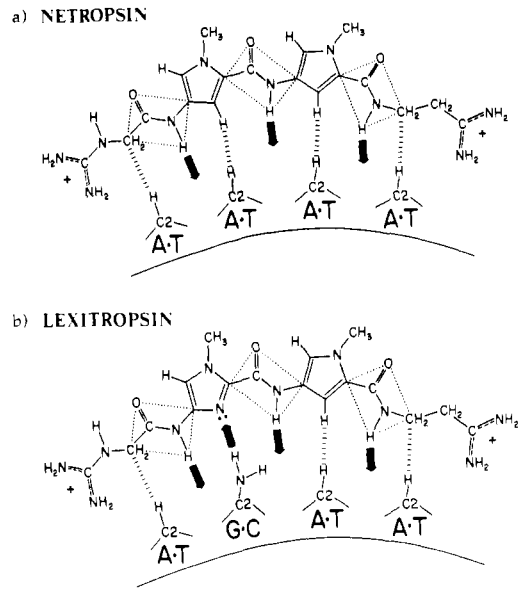

FIGURE 1: Schematic of the binding to the floor of the minor groove of (a) the naturally occurring drug netropsin and (b) its synthetic analogue lexitropsin. Black arrows are bifurcated hydrogen bonds to adenine N3 and/or thymine O2 atoms on the floor of the groove. In netropsin, A·T specificity is conferred because two methylene -CH_2- "elbows" and two pyrrole CH groups sit in close van der Waals packing contact with the C2 hydrogens of adenines (dashed lines), leaving no room for the bulkier C2 amines of guanines. Replacement of one pyrrole by an imidazole in lexitropsin should permit acceptance of a G·C base pair at that point, as drawn [from Kopka et al. (1985a,b)].

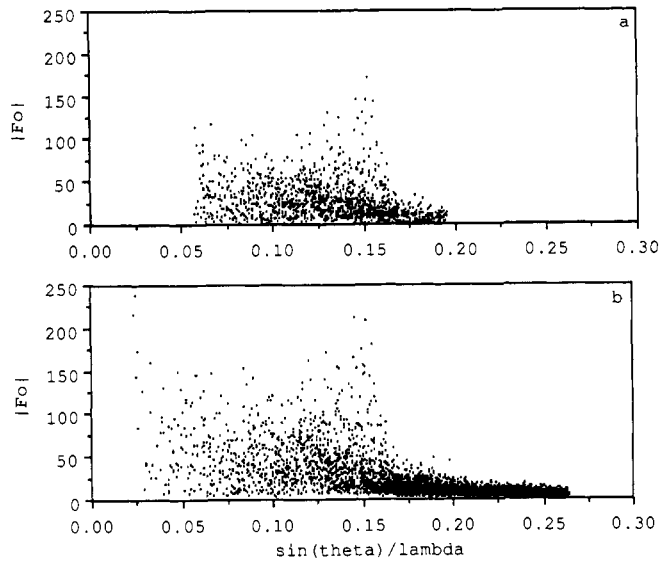

FIGURE 2: Scatter plots of decay-corrected structure-factor observations (F_o) vs resolution (sin θ/λ) for (a) Lex1 data and (b) Drew dodecamer data. The right-hand termination of the data in (a) is at $d = 1/2(\sin\theta/\lambda) = 2.5$ Å and in (b) is at $d = 1.9$ Å. In both cases, a precipitous fall-off of strong reflections is observed at around $d = 3$ Å.

instead of A·T. The C-G-T-G-A-A-T-T-C-A-C-G sequence was designed to contain two such sites, symmetrically disposed about the middle of the chain as shown by underlining. For brevity, C-G-C-G-A-A-T-T-C-G-C-G will be referred to as the "Drew" sequence and C-G-T-G-A-A-T-T-C-A-C-G as the "Lex1" sequence. The Lex1 sequence retains the central A-A-T-T netropsin-binding site of the Drew sequence. Hence if a Lown lexitropsin molecule *prefers* a site with one G·C base pair, it should bind to one of the two off-center sites of Lex1. However, footprinting experiments with the first synthetic lexitropsins (Lown et al., 1986b) showed them to be G·C accepting rather than G·C specific. Consequently, a lexitropsin may bind only to the central A-A-T-T netropsin site of this dodecamer. The ultimate goal remains to cocrystallize one of the Lown lexitropsins with this Lex1 sequence, but solution of the structure of the DNA alone has been undertaken as a preliminary step.

Materials and Methods

Synthesis and Crystal Growth. C-G-T-G-A-A-T-T-C-A-C-G was synthesized manually by the solid-phase phosphoramidite method (Sinha et al., 1984; Matteucci & Caruthers, 1981) and purified on a cellulose DE52 denaturing anion-exchange column as described previously (Wing et al., 1980). Crystals were grown by vapor diffusion at room temperature, in sitting drops containing 0.23 mM DNA, 3.8 mM $Mg(OAc)_2$, 0.8 mM spermine, and 13% 2-methyl-2,4-pentanediol (MPD). Crystals appeared after equilibration to 27.5% MPD. Isomorphous crystals also could be obtained by using equivalent concentrations of $Ca(OAc)_2$ or $CaCl_2$ in place of magnesium acetate. These conditions are similar to those used with the Drew dodecamer, except for an increase in the spermine to duplex ratio from 0.8:1 to 3:1.

Data Collection. Intensity data were collected on Rigaku AFC5R diffractometer. Many crystals of quite variable quality were mounted for data collection. Some crystals gave sharp and narrow peak profiles, while most gave broader profiles or were even twinned. At room temperature the diffraction pattern decayed within a day. Cooling improved crystal stability; several data sets were collected successfully at 0 °C. Attempts at collecting data at −25 °C and −100 °C were unsuccessful; the crystals either gave no diffraction pattern or decayed too rapidly for data collection. In all cases, the observed intensities fell off rapidly with resolution.

The data set selected for the analyses presented in this report is strong and extends to a 2.5-Å resolution. Figure 2 compares scatter plots of F_o against resolution for the data set used here and for the Drew dodecamer data. The data are quite strong out to a resolution of $d = 3$ Å or sin $\theta/\lambda = 0.167$ and then continue with significant F_o's to the 2.5-Å cutoff. The Drew dodecamer data also show the 3.0-Å boundary, with additional weaker reflections beyond 2.5 Å.

All crystals mounted exhibited significant mosaic spread, suggesting internal crystalline disorder. This disorder could be dynamic (arising from thermal vibrations), or static (arising from microheterogeneity of individual molecules in the crystal lattice), or some combination of the two factors. At intermediate resolution, ca. 2.5 Å, the limitations imposed by disorder must always be kept in mind.

The data set used for refinement was collected from a crystal of approximate dimensions of 0.15 × 0.20 × 1.2 mm, with a strong pattern and reliable counting statistics out to a resolution of 2.5 Å. This resolution limit was imposed by time-dependent crystal decay. The data are isomorphous with those of the Drew dodecamer (Wing et al., 1980), in the orthorhombic space group $P2_12_12_1$ and with cell dimensions as listed in Table I.

NUCLSQ *Refinement.* Coordinates from the Drew dodecamer were modified to convert cytosines 3 and 15 to thymines and to convert guanines 10 and 22 to adenines. These new coordinates then were used as a starting point for refinement of the Lex1 structure with the NUCLSQ refinement package of Westhof et al. (1985).

The starting Lex1 coordinates gave a residual error of $R = 38.4\%$, a reasonable beginning. Round 1 of refinement with NUCLSQ applied positional minimization to the 3.0-Å data and lowered the R factor to 32.0%. Round 2 extended the data

Table I: Data Collection Statistics for the Lex1 and Drew Dodecamers[a]

data set	cell dimen. (Å) a	b	c	resolution range	no. data collected	no. F_o's nonzero	above 2σ
Lex1	24.94	40.78	66.13	10–2.5 Å	1835	1475	
Drew	24.87	40.39	66.20	10–1.9 Å			2725

[a] Space group $P2_12_12_1$, one double helix per asymmetric unit.

Table II: R Factor and Model Comparisons for Preliminary Refinement

	starting models unshifted		shifted	no. reflect. used
initial R factor	38.4%		34.8%	1255
rms model difference		0.39 Å		
round 1 R factor[a]	32.0%		27.1%	1255
round 3 R factor[b]	20.9%		20.1%	1475
rms model difference		0.19 Å		

[a] Round 1 applied positional refinement to the 3.0-Å data. [b] Round 3 applied positional and temperature-factor refinement to all of the data.

Table III: Final Map Characteristics

resolution range (Å)	10–2.5	σ level on Fourier maps (e/Å^3)	0.19
no. nonzero reflections	1475	σ level on difference maps (e/Å^3)	0.030
final R factor	15.8%	highest difference peak (e/Å^3)	0.14
rms bond lengths (Å)	0.020	lowest difference peak (e/Å^3)	−0.14
no. solvent peaks	36	lowest peak assigned (e/Å^3)	0.09

set to 2.5 Å, and round 3 added temperature-factor refinement, yielding a residual error of R = 20.9%.

As a procedural check, the entire refinement was repeated with the starting Lex1 helix first subjected to rigid-body refinement by using XPLOR (Brünger et al., 1987). This positional minimization reduced the residual error by 3.6% to R = 34.8% (Table II). The unshifted and shifted starting models showed a root-mean-square (rms) difference in atomic positions of 0.39 Å before NUCLSQ refinement, a significant difference. Subsequent refinement of the shifted starting model lowered the residual error to 20.1% and brought the two refined models to a rms difference of only 0.19 Å. The two refinements were conducted absolutely in parallel so that their end products would be comparable. The two models were considered to be essentially identical after round 3, and refinement was completed by using the shifted model.

Solvent Peak Assignment. Searches for solvent peaks were carried out with Fourier and difference Fourier maps (Table III). At each round of analysis, approximately 10 solvent peaks that were greater than three times the sigma value of the ($F_o - F_c$) difference map were added. This minimum peak height ranged from 0.2e/Å^3 to 0.1e/Å^3, depending on the stage of addition of solvent atoms to the structure. Difference map peaks lower than 0.1e/Å^3 were felt to be too far below the noise level of the maps to justify their interpretation as solvent.

The peaks selected were inspected by using FRODO on an Evans and Sutherland PS390. A solvent atom was deleted if it was closer than 2.0 Å to the DNA or to another solvent atom. Atoms farther than 4.0 Å from the DNA or from other solvent atoms also were deleted, with the expectation that if they were real they would be recovered at a later cycle of solvent examination, after the undershell of solvent had been established. Repeated cycles of map calculation and inspection, and solvent assignment, were conducted until all difference peaks above 0.10e/Å^3 either were assigned or were rejected because of unfavored contacts.

Temperature-Factor Trends. The overall mean value of the thermal parameter B of 15.1 Å^2 is within an acceptable range for macromolecular structures. As with other dodecamer

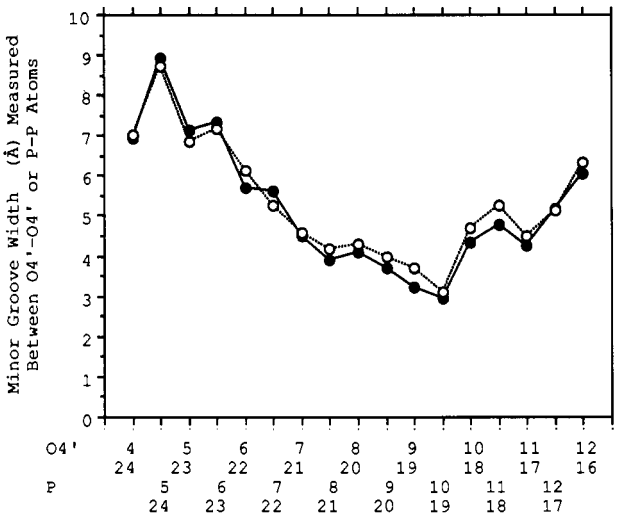

FIGURE 3: Plot of the shortest width across the minor groove, as measured between sugar O4′ atoms or phosphate P atoms. The O–O distances have been reduced by 2.3 Å to represent two nonbonded van der Waals oxygen radii, and P–P distances have been reduced by 5.8 Å to represent two phosphate group radii. Hence the plotted distance in each case is the width of the gap between atoms on opposite walls. The filled circles and solid line represent Lex1. The open circles and dotted line represent the parent dodecamer.

structures, mean temperature factors increase from bases to sugars to phosphates, as though the entire double helix were subject to a certain amount of librational disorder about its axis. The B values of a few of the base atoms in both data sets refined to zero, a physically meaningless situation that merely reflects the inherent uncertainty in refined B values. These were checked by setting the offending B values to the mean for the entire model and allowing them to refine once again to an extremely low value.

The data ultimately refined to a residual error of 15.8% for 1475 nonzero reflections out to a resolution of 2.5 Å, with 36 waters.

Results and Discussion

Hydration Comparisons. The observable extent of hydration of Lex1 differs significantly from the parent structure, a feature that probably is to be ascribed to the more limited resolution of the Lex1 data (Figure 3). Yet of the 36 waters localized, 30 lie within 2 Å of a corresponding one of the 80 waters found in the parent Drew dodecamer, C-G-C-G-A-A-T-T-C-G-C-G. Evidently the Lex1 structure exhibits an incomplete sampling of the same set of water positions that is seen in greater completeness in the Drew structure. This, in turn, lends credibility to the water positions that were assigned in the Drew structure a decade ago (Drew & Dickerson, 1981).

The model shows a fragmentary spine of hydration in the minor groove. Figure 4 illustrates the complete spine found in the Drew helix and those portions that are clear in the Lex1 model. Other structure analyses with a comparable intermediate resolution of d = 2.3–2.7 Å and with data collected at room temperature or 4 °C have exhibited comparable fragmentary spines in the central narrow minor-groove region (Kopka et al., 1983; Nelson et al., 1987; Yoon et al., 1988).

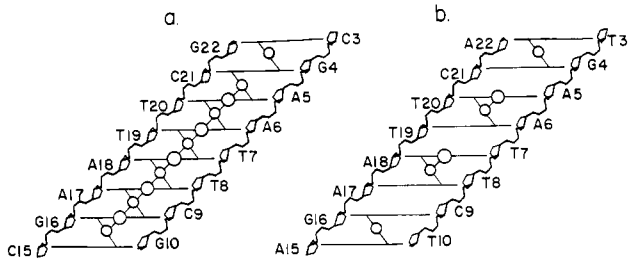

FIGURE 4: Unrolled-helix schematics of the minor groove, showing the observed minor-groove hydration for (a) the Drew dodecamer and (b) Lex1. The fragmentary nature of the spine of hydration in (b) could arise from the lower resolution of the analysis, from as yet undetermined sequence effects, or from the temperature at which data were collected. Narendra et al. (1991) report a more well developed spine with the same DNA sequence for data collected at low temperature.

Table IV: Selected Helical Parameters for Model Comparisons

parameter	CGCGAATTCGCG	CGTGA-ATTCACG
propeller	−12 ± 6[a]	−14 ± 5
buckle	0.4 ± 5	0.4 ± 5
twist	36 ± 4	36 ± 4
rise per base pair	3.3 ± 0.1	3.3 ± 0.4
residues per turn	10.0	9.9
tilt	−0.1 ± 2	−0.4 ± 2
roll	0.3 ± 5	0.6 ± 4
bend[b]	12	10

[a] Mean ± standard deviation. [b] The bend is calculated as the difference between the averaged four base normals at each end of the helix.

But Narendra et al. (1991) observe a nearly complete minor-groove spine in a 2.7-Å analysis of the Lex1 sequence, C-G-T-G-A-A-T-T-C-A-C-G, with data collected at −90 °C. The failure to observe a spine of hydration in some room-temperature or 4 °C dodecamer structures seems to be a consequence of the low resolution of the data, but this can be counteracted by immobilizing the solvent at very low temperature. The new structure provides us with a single "snapshot" of part of the complete spine. Repeated collection of other data sets under equivalent conditions might yield other partial snapshots, but one could never be sure whether a peak that appeared in only one analysis was a valid fragment of the complete water structure or was only an artifact of that analysis that should be discarded.

Comparisons of Lex1 and Other Structures. The Drew C-G-C-G-A-A-T-T-C-G-C-G and Lex1 C-G-T-G-A-A-T-T-C-A-C-G helices, compared in Figure 5, are virtually identical. As has been mentioned, minor-groove width is unchanged by the introduction of two new A·T base pairs at the positions underlined in the sequence just given. The overall bend in helix axis (Table IV) is comparable for the Drew and Lex1 helices and indeed for every other dodecamer structure except C-G-C-G-A-A-T-T-^{Br}C-G-C-G, where the major-groove bromines on the cytosines inhibit bending (Fratini et al., 1982). Table V shows, however, that the *propeller* of the new A·T pairs is larger in magnitude than that of the G·C pairs that were replaced, in agreement with the general observation that A·T pairs tend to have a greater propeller value.

The Lex1 sequence provides a new example of a C-A base step, which appears to behave differently in dodecamers and decamers. The twist of a C-A step is smaller than 36° in all of the orthorhombic $P2_12_12_1$ dodecamers:

C-G-T-G-A-A-T-T-C-A-C-G	30°, 30°	Lex1
C-G-C-A-T-A-T-A-T-G-C-G	34°, 33°	Yoon et al. (1988)
C-G-C-A-A-A-T-T-T-G-C-G	27°, 34°	Coll et al. (1987)
C-G-C-A-A-A-A-A-A-A-G-C-G	29°,	Nelson et al. (1987)

Table V: Propeller Values, Means, and Standard Deviations[a]

base pair	CGCGAATTCGCG	CGTGAATTCACG
C1/G24	−13	−11
G2/C23	−11	−13
Y3/R22	−4	−11
G4/C21	−10	−12
A5/T20	−16	−12
A6/T19	−18	−19
T7/A18	−17	−21
T8/A17	−17	−17
C9/G16	−16	−19
R10/Y15	−5	−15
C11/G14	−17	−16
G12/C13	−2	−4
mean	−12	−14
SD	6	5

[a] Propeller values, means, and standard deviations for the parent sequence (CGCGAATTCGCG) and Lex1 (CGTGAATTCACG). Base-pair steps are given in the left-most column except for where the sequences differ, at the 3/10 and 10/15 steps, which are indicated by the symbols R = purine and Y = pyrimidine. The propeller value for these steps in Lex1 are underlined to highlight this difference between the models.

But is is considerably larger than 36° in each of the four presently known decamer sequences in which it occurs, in the monoclinic space group $C2$, or in the hexagonal space group $P6$:

C-C-A-A-G-A-T-T-G-G	C2	50°	Privé et al. (1987)
C-C-A-A-C-G-T-T-G-G	C2	51°	Privé et al. (1990)
C-C-A-G-G-C-C-T-G-G	C2	51°	Heinemann and Alings (1989)
C-C-A-G-G-C-^{me}C-T-G-G	P6	44°, 48°	Heinemann and Alings (1991)

What is the determining difference between these two classes? The facile but incorrect answer is "crystal packing". All of the known dodecamers are isomorphous and have identical intermolecular contacts within the crystal; and three of the decamers are themselves mutually isomorphous with their own unique contacts. Could the C-A step be bistable (or even multistable), so that the dodecamer environment pushes it to a low twist, whereas the decamer environment causes a high twist? It is not easy, even from close inspection of stereopairs showing intermolecular crystal packing, to argue the inevitability of crystal packing as an explanation (Dickerson et al., 1987; Privé et al., 1990; Yanagi et al., 1990). And the recent finding of Heinemann and Alings (1991) that their methylated sequence adopts a quite different hexagonal packing but exhibits almost identical helix parameters with the unmethylated form gives strong support to the idea that these parameters are dictated mainly by the base sequence and not by crystal packing.

Another factor governing twist at a C-A step may be the sequence context in which that step is found. It is well-known that there are 10 unique steps of two successive base pairs, if self-complementarity of steps such as C-A and T-G are taken into account. It is less well-known that there are 136 different four-base steps with the same corrections for self-complementarity (Yanagi et al., 1990). Each dimer that is not itself self-complementary has 16 possible flanking sequences, and each self-complementary dimer has 10 different flanking sequences, with a numerology of $6 \times 16 + 4 \times 10 = 136$. Hence there are 16 different sequence contexts in which a C-A step can be located. Is this the factor that differentiates between high and low twist at C-A steps? Do these context-dictated twist steps then merely choose the crystal form with which they are most compatible? To put the dilemma succinctly, is the dodecamer C-A twist angle 29° because it is made so by the

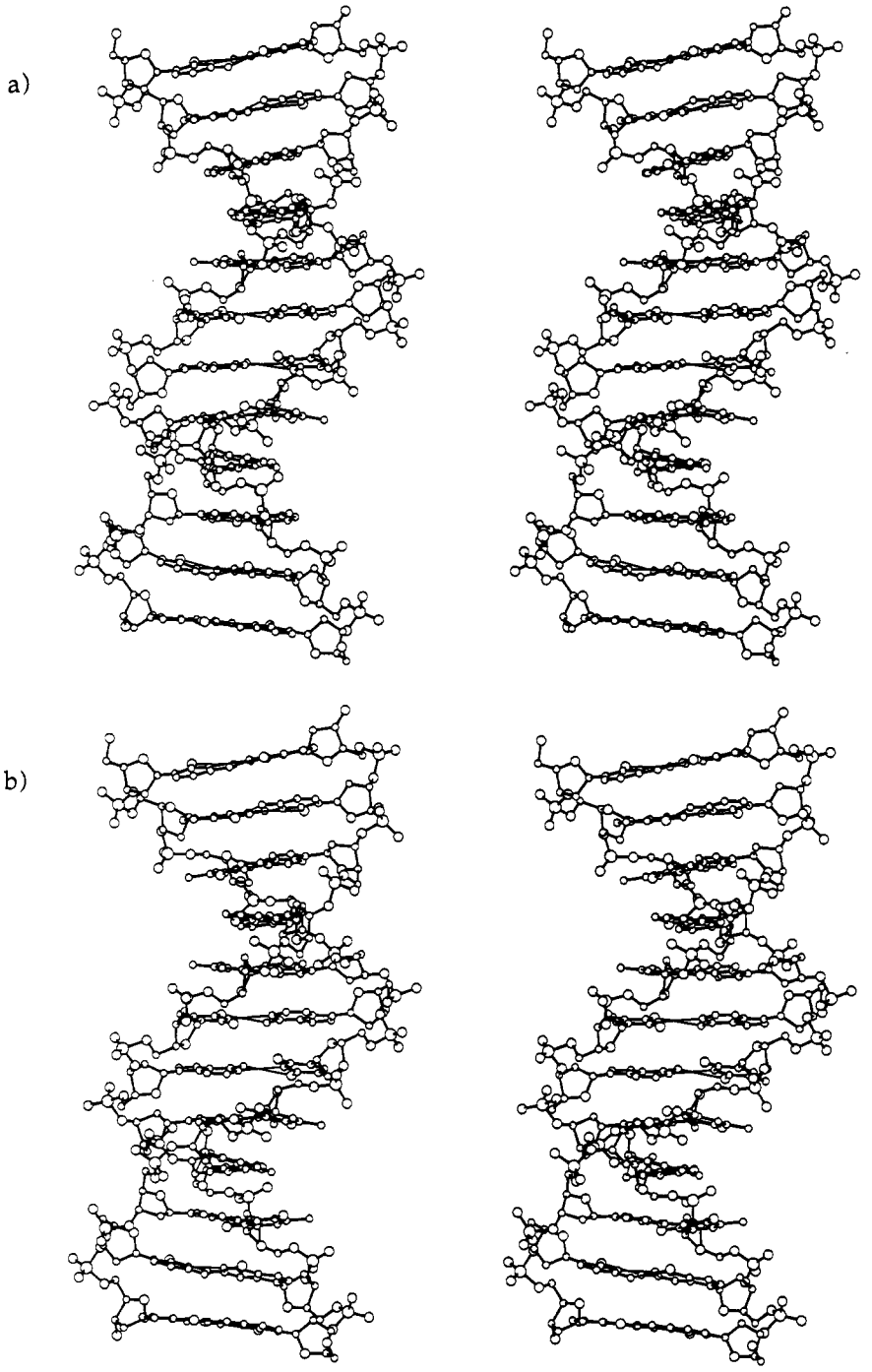

FIGURE 5: Stereopair drawings of (a) the Drew dodecamer C-G-C-G-A-A-T-T-C-G-C-G and (b) Lex1, C-G-T-G-A-A-T-T-C-A-C-G.

orthorhombic $P2_12_12_1$ crystal form, or does the helix adopt this form because of a preexisting 29° twist at its C-A step? A parallel query can be framed about the 50° C-A twist in the decamers.

At present we do not have enough data to decide this issue. The large twist only occurs at steps of the type Y-C-A-R, where Y = pyrimidine (C, T) and R = purine (G, A). Is this a meaningful observation, or is it just a reflection of the lack of information? At present, even with 35 published structures of B-DNA oligomers, there are no examples of Y-C-A-R in any dodecamer. All that can be said is that the Y-C-A-R hypothesis has a certain credibility and is easily tested when we have more information.

The Lex1 crystals are isomorphous with those of the Drew dodecamer but show significantly less order and a poorer diffraction pattern. Could this arise from the simple substitution of the third base pair in from each end of the helix? Figure 4 of Dickerson et al. (1985) and Figure 9 of Dickerson and Drew (1981) exhibit stereopair drawings of the overlap between helices in the Drew dodecamer, viewed in two perpendicular directions. The helices overlap by two base pairs: the final base pair on one helix interacts with the penultimate base pair on the other and vice versa. The interaction consists of a face-to-face contact of minor grooves and involves only guanines: one guanine donates a N2 amine hydrogen bond to the guanine N3 on the other helix and receives an equivalent hydrogen bond at its own N3. The third base pair in from each end appears to be only peripherally involved with the neighboring helix. The 3′-OH of strand 1 is hydrogen bonded to the N2 amino group of G22 at the third base pair of a neighboring helix, but the 3′-OH of strand 2 is bonded not to the symmetry-equivalent N2 of G10 but to the N3 of G16 on the fourth base pair in from the end. Exactly the same hydrogen bonds are formed in the C-G-C-G-A-A-T-T-^{Br}C-G-

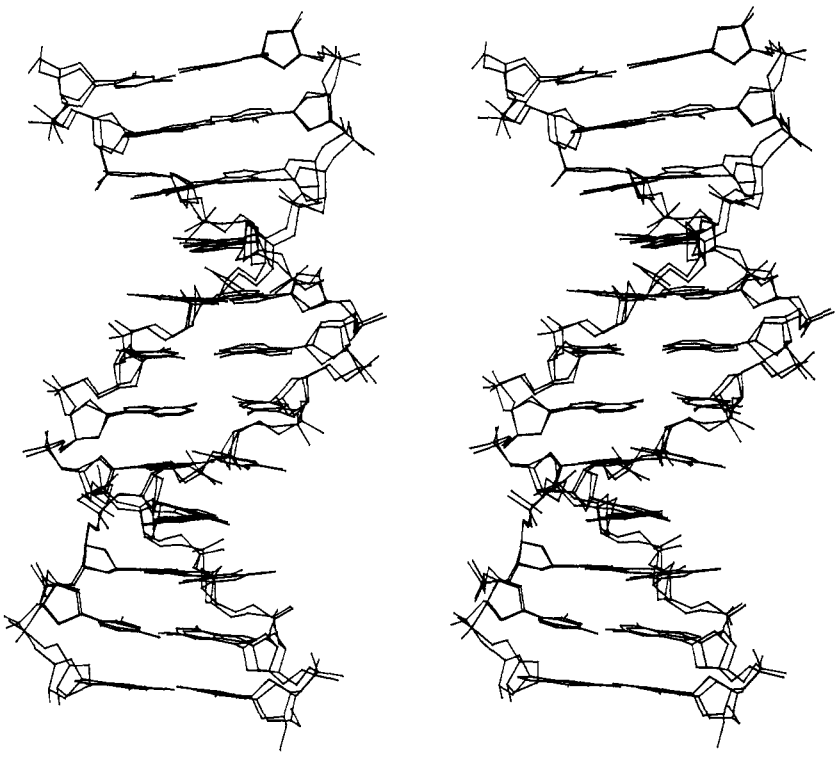

FIGURE 6: Superimposed stereopairs of the two independently-solved structures for C-G-T-G-A-A-T-T-C-A-C-G, from the present work and from Narendra et al. (1991).

C-G or MPD7 structure, in which the end in helix axis has been removed by major-groove bromines [Figure 2 of Dickerson et al. (1987)].

The Lex1 structure shows the same pattern of N2···N3 and N3···N2 bonds between the first two base pairs. The bond involving the 3′-OH of strand 2 remains as in the Drew structure, but the bond from the 3′-OH of strand 1 to the third base pair is lost because of the sequence substitution. Instead, a van der Waals contact is made with the adenine-ring C2 atom. Hence the packing is preserved even though one hydrogen bond is lost. It seems unlikely that the loss of one hydrogen bond out of a set of six could be solely responsible for the observed loss of order in the Lex1 crystals.

Concluding Remarks

The overlap of minor grooves at the ends of neighboring helices has one unfortunate consequence for lexitropsin binding. If a lexitropsin were to bind preferentially to one of its two target sites, underlined in C-G-T-G-A-A-T-T-C-A-C-G, it would prevent adoption of the particular packing scheme found with DNA alone because it would block a region where the minor grooves of neighboring molecules overlap. Hence if we can obtain crystals of a Lex1–lexitropsin complex, we can anticipate their adoption of a new packing scheme.

Narendra et al. (1991) have solved the same C-G-T-G-A-A-T-T-C-A-C-G structure at −90 °C, and their laboratory has exchanged data and coordinates with ours. The two structures appear to be similar overall but with differences in detail. Figure 6 is a superposition of the two structures: model one of this report and the final refined coordinates of Narendra et al. The two laboratories have entered into a collaboration to address several issues raised here, including those of data quality, solvent structure, and crystal polymorphism, as well as considerations of differences in crystallization conditions, data collection, and refinement strategies.

Acknowledgments

We would like especially to thank Prof. Helen Berman and her colleagues for their generous spirit, which turned what could have been a competition into a fruitful collaboration. We also appreciate the technical help of Maria K. Grzeskowiak in growing crystals.

Registry No. d(C-G-T-G-A-A-T-T-C-A-C-G), 113725-84-1; d(G-C-G-A-A-T-T-C-G-C-G), 77889-82-8; lexitropsin, 121854-21-5.

References

Brünger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458–460.

Coll, M., Frederick, C. A., Wang, A. H.-J., & Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385–8389.

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761–786.

Dickerson, R. E., Kopka, M. L., & Pjura, P. (1985) in *Biological Macromolecules and Assemblies: Vol. 2—Nucleic Acids and Interactive Proteins* (Jurnak, F., & McPherson, A., Eds.) pp 37–127, Wiley, New York.

Dickerson, R. E., Kopka, M. L., & Pjura, P. (1986) *Chem. Scr. 26B*, 139–145.

Dickerson, R. E., Goodsell, D. S., Kopka, M. L., & Pjura, P. E. (1987) *J. Biomol. Struct. Dyn. 5*, 557–579.

Drew, H. R., & Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535–556.

Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686–14707.

Heinemann, U., & Alings, C. (1989) *J. Mol. Biol. 210*, 369–381.

Heinemann, U., & Alings, C. (1991) *EMBO J.* (in press).

Kopka, M. L., Fratini, A. V., Drew, H. R., & Dickerson, R. E. (1983) *J. Mol. Biol. 163*, 129–146.

Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985a) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1376–1380.

Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985b) *J. Mol. Biol. 183*, 553–563.
Lown, J. W., & Krowicki, K. (1985) *J. Org. Chem. 50*, 3775–3779.
Lown, J. W., Krowicki, K., Balzarini, J., & De Clercq, E. (1986a) *J. Med. Chem. 29*, 1210–1214.
Lown, J. W., Krowicki, K., Bhat, U. G., Skorobogaty, A., Ward, B., & Dabrowiak, J. C. (1986b) *Biochemistry 25*, 7408–7416.
Matteucci, M. D., & Caruthers, M. H. (1981) *J. Am. Chem. Soc. 103*, 3185–3191.
Narendra, N., Ginell, S. L., Russu, I. M., & Berman, H. B. (1991) *Biochemistry* (following paper in this issue).
Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature 330*, 221–226.
Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L. S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498–504.
Privé, G. G., Yanagi, K., & Dickerson, R. E. (1990) *J. Mol. Biol. 216*, (in press).
Pullman, B. (1983) *J. Biomol. Struct. Dyn. 1*, 773–794.
Sinha, N. D., Biernat, J., McNanus, J., & Koster, H. (1984) *Nucleic Acids Res. 12*, 4539–4557.
Wade, W. S., & Dervan, P. B. (1987) *J. Am. Chem. Soc. 109*, 1574–1575.
Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.
Wing, R. M., Drew, H. R., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature 287*, 755–758.
Yanagi, K., Privé, G. G., & Dickerson, R. E. (1990) *J. Mol. Biol. 216*, (in press).
Yoon, C., Privé, G. G., Goodsell, D. S., & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332–6336.
Zimmer, C. (1975) *Prog. Nucleic Acid Res. Mol. Biol. 15*, 285–318.

Crystal and Molecular Structure of a DNA Fragment: d(CGTGAATTCACG)[†,‡]

Narendra Narayana,[§,∥] Stephan L. Ginell,[§] Irina M. Russu,[⊥] and Helen M. Berman*[,§]
Department of Chemistry, Rutgers University, Piscataway, New Jersey 08855, and Department of Molecular Biology and Biochemistry, Wesleyan University, Middletown, Connecticut 06457

Received September 11, 1990; Revised Manuscript Received January 2, 1991

ABSTRACT: The crystal structure of the dodecanucleotide d(CGTGAATTCACG) has been determined to a resolution of 2.7 Å and refined to an *R* factor of 17.0% for 1532 reflections. The sequence crystallizes as a B-form double helix, with Watson–Crick base pairing. This sequence contains the *Eco*RI restriction endonuclease recognition site, GAATTC, and is flanked by CGT on the 5′-end and ACG on the 3′-end, in contrast to the CGC on the 5′-end and GCG on the 3′-end in the parent dodecamer d-(CGCGAATTCGCG). A comparison with the isomorphous parent compound shows that any changes in the structure induced by the change in the sequence in the flanking region are highly localized. The global conformation of the duplex is conserved. The overall bend in the helix is 10°. The average helical twist values for the present and the parent structures are 36.5° and 36.4°, respectively, corresponding to 10 base pairs per turn. The buckle at the substituted sites are significantly different from those seen at the corresponding positions in the parent dodecamer. Step 2 (GpT) is underwound with respect to the parent structure (27° vs 36°) and step 3 (TpG) is overwound (34° vs 27°). There is a spine of hydration in the narrow minor groove. The N3 atom of adenine on the substituted A10 and A22 bases are involved in the formation of hydrogen bonds with other duplexes or with water; the N3 atom of guanine on G10 and G22 bases in the parent structure does not form hydrogen bonds.

The *Eco*RI restriction endonuclease specifically recognizes the double-stranded DNA sequence 5′-GAATTC-3′ and cleaves it at the GpA bonds. The catalytic rates of the enzyme can be modulated by the sequences flanking the *Eco*RI recognition site. This flanking-sequence effect was first recognized by Halford et al. (1980). These authors studied the reaction of *Eco*RI endonuclease on λ phage DNA, which contains five *Eco*RI sites, numbered srI 1 to srI 5 (Daniels et al., 1983). The reaction rate at site srI 5 (5′...-TGAATTCA-...3′) was found to be enhanced relative to that at site srI 2 (5′...-TGAATTCT...3′). Moreover, the cleavage rates at these two sites exhibited markedly different dependencies on ionic strength and pH. In contrast, the cleavage rate at site srI 4 (5′...-AGAATTCT...3′) was the same as that at site srI 5. These results suggest that the orientation of the AT base pairs adjacent to the *Eco*RI site affects the recognition of this site by the protein. Specifically, when the orientation of the AT base pairs flanking the *Eco*RI site preserves the palindromic character of the site, the cleavage rates are increased (e.g., srI 4 and srI 5 on λ DNA). Otherwise, the cleavage rates are lower (e.g., srI 2 on λ phage).

Biochemical evidence indicates that the contacts of the protein with bases outside the recognition site play an important role in binding and cleavage. Strong ethylation interference has been observed for the two phosphate groups upstream from the recognition site (i.e., ...pNpGAATTC...) (Lu et al., 1981; Lesser et al., 1990). The patterns of ethy-

[†] This work was supported by grants from the NIH (GM 21589) for H.M.B. and the NSF (88-17589) for I.M.R.
* Address correspondence to this author.
[‡] The coordinates as well as the structure factors were deposited in the Protein Data Bank and the Nucleic Acid Database.
[§] Rutgers University.
[∥] Present address: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511.
[⊥] Wesleyan University.

0006-2960/91/0430-4449$02.50/0